JOURNAL OF MULTIVARIATE ANALYSIS 12, 346–370 (1982)

Asymptotic Properties of Projections with Applications to Stochastic Regression Problems

T. L. LAI* AND C. Z. WEI†

Columbia University and University of Maryland

Communicated by E. J. Hannan

Almost sure convergence properties of least-squares estimates in stochastic regression models and an asymptotic theory of related Euclidean projections are developed herein. Applications to autoregressive processes and to dynamic input–output systems are also discussed.

1. INTRODUCTION

Consider the multiple regression model

$$y_n = \beta_1 x_{n1} + \cdots + \beta_p x_{np} + \varepsilon_n, \qquad n = 1, 2,..., \tag{1.1}$$

where the errors ε_n form a martingale difference sequence with respect to an increasing sequence of σ-fields $\mathscr{F}_n$ (i.e., ε_n is $\mathscr{F}_n$-measurable and $E(\varepsilon_n | \mathscr{F}_{n-1}) = 0$ for every n), and the design levels $x_{n1},..., x_{np}$ at stage n are $\mathscr{F}_{n-1}$-measurable random variables. Throughout the sequel we shall let X_n denote the design matrix $(x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$ and let $\beta = (\beta_1,..., \beta_p)'$, where the prime denotes transpose. If $X_n' X_n$ is nonsingular, then the least-squares estimate $b_n = (b_{n1},..., b_{np})'$ of β is given by

$$b_n = (X_n' X_n)^{-1} X_n' \mathbf{y}_n, \qquad \text{where } \mathbf{y}_n = (y_1,..., y_n)'. \tag{1.2}$$

Examples of stochastic regressors x_{ij} arise in time series models, dynamic input–output systems, adaptive stochastic approximation schemes, stochastic control and other applications. Motivated by these applications, in particular to recursive on-line identification of dynamic systems and to adaptive

Received August 3, 1981; revised October 26, 1981.
AMS 1980 subject classification: Primary 62J05, 62M10, 60F15. Secondary 93B30.
Key words and phrases: Stochastic regressors, least squares estimates, projections, strong consistency, dynamic models, autoregressive processes, minimum eigenvalue, martingales.

* Research supported by the National Science Foundation and the National Institute of Health.
† Research supported by the National Science Foundation.

0047-259X/82/030346–25$02.00/0

stochastic control problems, there has been considerable recent interest in the question of strong consistency of b_n in stochastic regression models, both in the statistical and in the engineering literature (cf. [1, 4, 6–8, 11–13]). Under minimal assumptions on the stochastic regressors x_{ij}, we have recently established in [10] the following result on the strong consistency of b_n.

THEOREM 1. *Suppose that in the regression model* (1.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$ such that*

$$\sup_n E(|\varepsilon_n|^\alpha | \mathscr{F}_{n-1}) < \infty \qquad \text{a.s. } \textit{for some } \alpha > 2. \tag{1.3}$$

Moreover, assume that the design levels $x_{n1},..., x_{np}$ at stage n are $\mathscr{F}_{n-1}$-measurable random variables for every n such that

$$\lambda_{\min}(X_n' X_n) \to \infty \quad \text{a.s.,} \quad \{\log \lambda_{\max}(X_n' X_n)\}/\lambda_{\min}(X_n' X_n) \to 0 \quad \text{a.s.} \tag{1.4}$$

Then $b_n \to \beta$ a.s.

In (1.4) and throughout the sequel, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the maximum and minimum eigenvalues of a matrix A. Condition (1.4) on the stochastic regressors is in some sense weakest possible, and an example in which the ratio in (1.4) converges a.s. to a finite nonzero limit but b_n fails to be strongly consistent is given in [10]. Theorem 1 improves the recent result of Anderson and Taylor [1] who assumed the condition $\lambda_{\max}(X_n' X_n) = O(\lambda_{\min}(X_n' X_n))$ a.s., and that of Christopeit and Helmes [4] who assumed the condition $(\lambda_{\max}(X_n' X_n))^r = O(\lambda_{\min}(X_n' X_n))$ a.s. for some $r > \frac{1}{2}$. Here and in the sequel, we use the notation $u_n = O(v_n)$ a.s. to denote that $\limsup_{n\to\infty} |u_n/v_n| < \infty$ a.s. Likewise, $u_n = o(v_n)$ a.s. means that $u_n/v_n \to 0$ a.s.

To apply Theorem 1 or the other aforementioned results in the literature, we have to estimate the order of magnitude of $\lambda_{\max}(X_n' X_n)$ and of $\lambda_{\min}(X_n' X_n)$. The former quantity can be easily estimated by the inequality

$$\lambda_{\max}(X_n' X_n) \leqslant \operatorname{tr}(X_n' X_n) = \sum_{j=1}^{p} \sum_{i=1}^{n} x_{ij}^2 \leqslant p \lambda_{\max}(X_n' X_n). \tag{1.5}$$

The analysis of $\lambda_{\min}(X_n' X_n)$, however, is often a much harder problem. In Section 3 we discuss a method of tackling this problem.

Let $c_j(X_n)$ denote the jth column vector of the matrix X_n and let $\hat{c}_j(X_n)$ denote the projection of $c_j(X_n)$ into the linear space spanned by the other column vectors of X_n. Thus, in particular, $\hat{c}_1(X_n)$ is the projection of $c_1(X_n)$ into the linear space spanned by $c_2(X_n),..., c_p(X_n)$. Let $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n u_i v_i$

denote the Euclidean inner product of $\mathbf{u} = (u_1, \dots, u_n)'$ and $\mathbf{v} = (v_1, \dots, v_n)'$, and let $\|\mathbf{u}\| = (\sum_{i=1}^n u_i^2)^{1/2}$. Obviously,

$$\begin{aligned} X_n' X_n \text{ is invertible} &\Leftrightarrow \lambda_{\min}(X_n' X_n) > 0 \\ &\Leftrightarrow c_j(X_n) - \hat{c}_j(X_n) \neq 0 \qquad \text{for all} \quad j = 1, \dots, p. \end{aligned}$$

As will be shown in Section 3, $\lambda_{\min}(X_n' X_n)$ is related to the residual sum of squares $\|c_j(X_n) - \hat{c}_j(X_n)\|^2$ through the inequalities

$$\begin{aligned} p \min_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\|^2 &\geqslant \lambda_{\min}(X_n' X_n) \\ &\geqslant p^{-1} \min_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\|^2. \end{aligned} \tag{1.6}$$

Section 2 develops a general theory on the asymptotic properties of Euclidean projections of certain random vectors into linear spaces spanned by other random vectors satisfying certain measurability assumptions. Several applications of these results are given in Section 3. Among them are estimates of the order of magnitude of $\|c_j(X_n) - \hat{c}_j(X_n)\|^2$ ($j = 1, \dots, p$) which, together with (1.6), in turn provide estimates of the order of magnitude of $\lambda_{\min}(X_n' X_n)$. Another application is the proof of strong consistency of the least-squares identification method for dynamic input–output systems and for nonstationary autoregressive processes.

The asymptotic theory of projections developed in Section 2 also provides the following refinement of Theorem 1. While Theorem 1 considers the strong consistency of the entire vector $b_n = (b_{n1}, \dots, b_{np})'$, the following theorem gives less restrictive but analogous conditions that would imply the strong consistency of b_{nj} for a particular j.

THEOREM 2. *Suppose that in the regression model* (1.1), $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to an increasing sequence of* σ-*fields* $\{\mathscr{F}_n\}$ *such that* (1.3) *holds, and the design levels* $x_{n1}, \dots, x_{np}$ *at stage* n *are* $\mathscr{F}_{n-1}$-*measurable random variables for every* n. *For* $j = 1, \dots, p$, *the least-squares estimate* b_{nj} *of* β_j *can be represented as*

$$b_{nj} = \langle c_j(X_n) - \hat{c}_j(X_n), \mathbf{y}_n \rangle / \|c_j(X_n) - \hat{c}_j(X_n)\|^2. \tag{1.7}$$

Moreover, if

$$\|c_j(X_n) - \hat{c}_j(X_n)\| \to \infty \qquad \text{a.s.}$$

and

$$\max_{k \neq j} \log^+ \left(\sum_{i=1}^n x_{ik}^2 \right) = o(\|c_j(X_n) - \hat{c}_j(X_n)\|^2) \qquad \text{a.s.,} \tag{1.8}$$

then $b_{nj} \to \beta_j$ *a.s. (The notation* $\log^+ x$ *denotes the positive part of* $\log x$, *i.e.,* $\log^+ x = \log x$ *for* $x > 1$ *and* $\log^+ x = 0$ *if* $x \leqslant 1$.*)*

The proof of Theorem 2 is given in Section 3. It is based on a sharp estimate, developed in Section 2, of the order of magnitude of $\langle c_j(X_n) - \hat{c}_j(X_n), \boldsymbol{\varepsilon}_n \rangle$, where $\boldsymbol{\varepsilon}_n = (\varepsilon_1, \ldots, \varepsilon_n)'$. Note that in Theorem 2 we have not assumed that $X_n' X_n$ is eventually nonsingular with probability 1, which is, however, assumed in Theorem 1 in the condition $\lambda_{\min}(X_n' X_n) \to \infty$ a.s. From (1.5) and (1.6), it follows that

$$\max_{1 \leqslant k \leqslant p} \sum_{i=1}^{n} x_{ik}^2 \leqslant p \lambda_{\max}(X_n' X_n),$$

$$\lambda_{\min}(X_n' X_n) \leqslant p \min_{1 \leqslant j \leqslant p} \| c_j(X_n) - \hat{c}_j(X_n) \|^2.$$

Hence, if condition (1.4) of Theorem 1 is satisfied, then condition (1.8) of Theorem 2 is also satisfied for every j. Therefore, Theorem 1 can be obtained as a corollary of Theorem 2.

2. Some Asymptotic Properties of Projections

If A, B are matrices each having n rows, we let $L(A)$ denote the linear space spanned by the column vectors of A and let $L(A, B)$ denote the linear space spanned by the column vectors of A and B. Throughout this section we let $\{\mathscr{F}_n\}$ be an increasing sequence of σ-fields, and let $\{\varepsilon_n\}$ be a martingale difference sequence with respect to $\{\mathscr{F}_n\}$. Moreover, for each $n \geqslant 1$, let $v_n, w_n, z_{n1}, \ldots, z_{np}$ be $\mathscr{F}_{n-1}$-measurable random variables. Let $Z_n = (z_{ij})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p}$, $\boldsymbol{\varepsilon}_n = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\mathbf{w}_n = (w_1, \ldots, w_n)'$, $\mathbf{v}_n = (v_1, \ldots, v_n)'$. Let $\hat{\boldsymbol{\varepsilon}}_n, \hat{\mathbf{w}}_n, \hat{\mathbf{v}}_n$ denote the projections of $\boldsymbol{\varepsilon}_n$, $\mathbf{w}_n$, $\mathbf{v}_n$ into $L(Z_n)$. The following three theorems summarize the main results of this section.

Theorem 3. (i) *Assume that*

$$\sup_n E(\varepsilon_n^2 | \mathscr{F}_{n-1}) < \infty \qquad \text{a.s.} \tag{2.1}$$

Then for every $\delta > 0$,

$$\| \hat{\boldsymbol{\varepsilon}}_n \|^2 = O\left(\left\{ \log^+ \left(\sum_{j=1}^{p} \sum_{i=1}^{n} z_{ij}^2 \right) \right\}^{(1+\delta)} \right) \qquad \text{a.s.} \tag{2.2}$$

(ii) *If* (2.1) *is replaced by the stronger assumption* (1.3), *then* (2.2) *also holds with* $\delta = 0$.

THEOREM 4. (i) *If* (2.1) *holds, then for every* $\delta > 0$,

$$\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n \rangle = \langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle = \langle \mathbf{w}_n, \boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n \rangle$$

$$= O\left(\|\mathbf{w}_n - \hat{\mathbf{w}}_n\| \left[\max \left\{ 1, \log^+ (\|\mathbf{w}_n - \hat{\mathbf{w}}_n\|), \log^+ \left(\sum_{j=1}^{p} \sum_{i=1}^{n} z_{ij}^2 \right) \right\} \right]^{(1+\delta)/2} \right) \quad \text{a.s.} \qquad (2.3)$$

(ii) *If* (2.1) *is replaced by the stronger assumption* (1.3), *then* (2.3) *also holds with* $\delta = 0$.

THEOREM 5. *Assume that* (1.3) *holds and that* $\liminf_{n\to\infty} E(\varepsilon_n^2 | \mathscr{F}_{n-1}) > 0$ a.s. *Let* $\hat{\hat{\mathbf{w}}}_n$ *denote the projection of* $\mathbf{w}_n$ *into* $L(Z_n, \mathbf{v}_n + \boldsymbol{\varepsilon}_n)$ *and let* $\mathbf{v}_n^*$ *denote the projection of* $\mathbf{v}_n$ *into* $L(Z_n, \mathbf{w}_n)$. *Suppose that*

$$\log^+ \left(\sum_{j=1}^{p} \sum_{i=1}^{n} z_{ij}^2 + \sum_{i=1}^{n} w_i^2 \right) = o(n) \quad \text{a.s.} \qquad (2.4)$$

Then

$$\|\boldsymbol{\varepsilon}_n\|^2 = O(n) \quad \text{a.s.,} \qquad \liminf_{n\to\infty} n^{-1} \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 > 0 \quad \text{a.s.} \qquad (2.5)$$

and

$$\|\mathbf{w}_n - \hat{\hat{\mathbf{w}}}_n\|^2 = \left\{ \frac{\|\mathbf{v}_n - \mathbf{v}_n^*\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2}{\|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2} + o(1) \right\} \|\mathbf{w}_n - \hat{\mathbf{w}}_n\|^2 \quad \text{a.s.} \qquad (2.6)$$

The special case $v_n = 0$ for all n (and therefore $\mathbf{v}_n = \hat{\mathbf{v}}_n = \mathbf{v}_n^* = \mathbf{0}$) in Theorem 5 says that the residual sum of squares obtained by projecting $\mathbf{w}_n$ into $L(Z_n, \boldsymbol{\varepsilon}_n)$ is a.s. asymptotically equivalent to the residual sum of squares obtained by projecting $\mathbf{w}_n$ into $L(Z_n)$. This kind of results plays an important role in the applications to autoregressive processes and dynamic systems in Section 3 where Theorem 5 enables us to estimate the residual sum of squares by a stepwise procedure based on successive projections.

The estimates of the asymptotic order of magnitude for $\|\hat{\boldsymbol{\varepsilon}}_n\|$ in Theorem 3 and for $\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle$ in Theorem 4 are in a sense best possible, as is shown by the following two simple examples.

EXAMPLE 1. Consider the special case $p = 1$ in Theorem 3(i) and assume that z_n $(=z_{n1})$ are nonrandom constants and $\varepsilon_1, \varepsilon_2, \ldots$ are independent random variables with zero means and $\sup_n E\varepsilon_n^2 < \infty$. Since $\hat{\boldsymbol{\varepsilon}}_n = \{(\sum_1^n z_i \varepsilon_i)/(\sum_1^n z_i^2)\}(z_1, \ldots, z_n)'$, the asymptotic behavior of $\|\hat{\boldsymbol{\varepsilon}}_n\|^2 =$

$(\sum_1^n z_i\varepsilon_i)^2/(\sum_1^n z_i^2)$ is related to that of the weighted sum $\sum_1^n z_i\varepsilon_i$. Assuming that $\sum_1^\infty z_i^2=\infty$, (2.2) reduces in this case to

$$\sum_1^n z_i\varepsilon_i = O\left(\left(\sum_1^n z_i^2\right)^{1/2}\left(\log\sum_1^n z_i^2\right)^{(1+\delta)/2}\right) \quad \text{a.s.} \tag{2.7}$$

for every $\delta>0$. The following special case shows that we cannot set $\delta=0$ in (2.7). Let $z_n=2^n$, and let ε_n be independent symmetric random variables such that $E\varepsilon_n^2=1$ for all n and

$$P[|\varepsilon_n|=n^{1/2}(\log n)^{1/2}]=1-P[|\varepsilon_n|\leqslant 2]\sim 1/(n\log n)$$

as $n\to\infty$. The Borel–Cantelli lemma then implies that

$$P[|\varepsilon_n|=n^{1/2}(\log n)^{1/2}\ \text{i.o.}]=1. \tag{2.8}$$

If (2.7) should hold with $\delta=0$, then

$$z_n\varepsilon_n = O\left(\left(\sum_1^n z_i^2\right)^{1/2}\left(\log\sum_1^n z_i^2\right)^{1/2}\right) \quad \text{a.s.} \tag{2.9}$$

Since $z_n=2^n$, (2.9) would in turn imply that $\varepsilon_n=O((\log 2^{2n})^{1/2})=O(n^{1/2})$ a.s., contradicting (2.8). Since Theorem 3 includes the asymptotic behavior of weighted sums of independent random variables as a special case, we cannot expect to get sharper results than the logarithmic order in (2.2) without more stringent assumptions on the matrix Z_n.

EXAMPLE 2. Consider the simple linear model

$$y_n=\beta_1+\beta_2 x_n+\varepsilon_n, \tag{2.10}$$

where $\varepsilon_1,\varepsilon_2,\ldots$ are independent normal random variables with zero means and unit variances, and the random variables x_n are defined inductively by

$$x_1=0, \qquad x_{n+1}=\bar{x}_n+\bar{\varepsilon}_n, \qquad n\geqslant 1. \tag{2.11}$$

We use the notation $\bar{a}_n$ to denote the arithmetic mean of n numbers $a_1,\ldots,a_n$. The random variable x_n is clearly $\mathscr{F}_{n-1}$-measurable, where $\mathscr{F}_n$ is the σ-field generated by $\varepsilon_1,\ldots,\varepsilon_n$. Letting $\hat{\boldsymbol{\varepsilon}}_n$ denote the projection of $\boldsymbol{\varepsilon}_n$ into the linear space spanned by the column vectors $(1,\ldots,1)'$ and $(x_1,\ldots,x_n)'$ of the design matrix as before, we obtain that

$$\|\hat{\boldsymbol{\varepsilon}}_n\|^2=n\bar{\varepsilon}_n^2+\left\{\sum_1^n(x_i-\bar{x}_n)\varepsilon_i\right\}^2\Big/\left\{\sum_1^n(x_i-\bar{x}_n)^2\right\}, \tag{2.12}$$

which corresponds to the standard analysis of variance decomposition for the sum of squares due to regression in this case. As shown in [8],

$$\sum_1^n (x_i - \bar{x}_n)\,\varepsilon_i \sim \sum_1^n (x_i - \bar{x}_n)^2 \sim \log n \qquad \text{a.s.} \tag{2.13}$$

Since $n\bar{\varepsilon}_n^2 = O(\log\log n)$ a.s. by the law of the iterated logarithm, (2.12) and (2.13) imply that

$$\|\hat{\boldsymbol{\varepsilon}}_n\|^2 \sim \log n \qquad \text{a.s.} \tag{2.14}$$

Since $\log(n + \sum_{i=1}^n x_i^2) \sim \log n$ a.s. (cf. [10]), (2.14) shows that the estimate of $\|\hat{\boldsymbol{\varepsilon}}_n\|^2$ in Theorem 3(ii) cannot be improved to a lower order of magnitude.

Let $\mathbf{w}_n = (x_1,..., x_n)'$ and let $\hat{\mathbf{w}}_n$ denote the projection of $\mathbf{w}_n$ into the linear space spanned by the vector $(1,..., 1)'$. Then $\mathbf{w}_n - \hat{\mathbf{w}}_n = (x_1 - \bar{x}_n,..., x_n - \bar{x}_n)'$ and therefore

$$\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle = \sum_1^n (x_i - \bar{x}_n)\,\varepsilon_i. \tag{2.15}$$

By (2.13) and (2.15), $\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle \sim \log n$ a.s.; moreover, $\|\mathbf{w}_n - \hat{\mathbf{w}}_n\|^2 = \sum_1^n (x_i - \bar{x}_n)^2 \sim \log n$ a.s. Since $\log(n + \sum_{i=1}^n x_i^2) \sim \log n$ a.s. (cf. [10]), we have therefore shown that with probability 1

$$\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle \sim \log n \sim \|\mathbf{w}_n - \hat{\mathbf{w}}_n\| \left\{ \log \left(n + \sum_{i=1}^n x_i^2 \right) \right\}^{1/2},$$

showing that the estimate (2.3) with $\delta = 0$ in Theorem 4(ii) is sharp.

We prove Theorems 3 and 5 by applying Theorem 4. For the proof of Theorem 4, we make use of the following two lemmas. The first lemma is recently established in [10, Corollary 1]. The second lemma follows from the a.s. convergence of the martingale transform $\sum_1^n w_i \varepsilon_i$ in the event $\{\sum_1^\infty w_i^2 < \infty\}$ and the strong law for $\sum_1^n w_i \varepsilon_i$ in $\{\sum_1^\infty w_i^2 = \infty\}$ (cf. [10, Lemma 2(iii) and Corollary 2]).

LEMMA 1. *Let $\{\varepsilon_n\}$ be a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathscr{F}_n\}$ such that* (2.1) *holds. Let $T_n = (t_{n1},..., t_{np})'$ be an $\mathscr{F}_{n-1}$-measurable random vector for every n. Let V_n be the Moore–Penrose generalized inverse of $\sum_{i=1}^n T_i T_i'$ and let $N = \inf\{n: \sum_1^n T_i T_i'$ is nonsingular$\}$* $(\inf \phi = \infty)$.

(i) *In the event $\{N < \infty$ and $\lim_{n\to\infty} \lambda_{\max}(\sum_1^n T_i T_i') < \infty\}$,*

$$\sum_{k=N+1}^\infty \left(T_k' V_{k-1} \sum_{i=1}^{k-1} T_i \varepsilon_i \right)^2 \Big/ (1 + T_k' V_{k-1} T_k) < \infty \qquad \text{a.s.} \tag{2.16}$$

(ii) *In the event* $\{N < \infty, \lim_{n\to\infty} \lambda_{\max}(\sum_1^n T_i T_i') = \infty\}$,

$$\sum_{k=N+1}^{n} \left(T_k' V_{k-1} \sum_{i=1}^{k-1} T_i \varepsilon_i\right)^2 / (1 + T_k' V_{k-1} T_k)$$
$$= O\left(\left[\log \lambda_{\max}\left(\sum_1^n T_i T_i'\right)\right]^{1+\delta}\right) \quad \text{a.s.} \tag{2.17}$$

for every $\delta > 0$.

(iii) *If* (2.1) *is replaced by the stronger assumption* (1.3), *then* (2.17) *also holds with* $\delta = 0$ *in the event* $\{N < \infty, \lim_{n\to\infty} \lambda_{\max}(\sum_1^n T_i T_i') = \infty\}$.

LEMMA 2. *Let* $\{\varepsilon_n\}$ *be a martingale difference sequence with respect to an increasing sequence of* σ*-fields* $\{\mathscr{F}_n\}$. *Let* w_n *be an* $\mathscr{F}_{n-1}$*-measurable random variable for every* $n \geqslant 1$.

(i) *If* (2.1) *holds, then for every* $\delta > 0$

$$\sum_1^n w_i \varepsilon_i = O\left(\left(\sum_1^n w_i^2\right)^{1/2} \max\left\{1, \left(\log^+ \sum_1^n w_i^2\right)^{(1+\delta)/2}\right\}\right) \quad \text{a.s.} \tag{2.18}$$

(ii) *If* (1.3) *holds, then* (2.18) *holds with* $\delta = 0$.

Combining Lemma 1 with Lemma 2, we obtain the following result from which Theorem 4 follows.

LEMMA 3. *With the same notations and assumptions as in Lemma* 1, *let* $\{w_n\}$ *be a sequence of random variables such that* w_n *is* $\mathscr{F}_{n-1}$*-measurable for every* $n \geqslant 1$. *Let* $W_n = \sum_{i=1}^n w_i T_i$ *and let* $s_n = \sum_1^n (w_i - W_n' V_n T_i)^2$. *Then in the event* $\{N < \infty\}$, *we have for every* $\delta > 0$

$$\sum_1^n (w_i - W_n' V_n T_i)\, \varepsilon_i$$
$$= O\left(s_n^{1/2}\left[\max\left\{1, \log^+ s_n, \log^+\left(\lambda_{\max}\left(\sum_1^n T_i T_i'\right)\right)\right\}\right]^{(1+\delta)/2}\right) \quad \text{a.s.} \tag{2.19}$$

Moreover, if (1.3) *holds, then* (2.19) *also holds with* $\delta = 0$ *in the event* $\{N < \infty\}$.

Proof. Let $d_n = w_n - W_n' V_n T_n$. By Lemma 3 of [9], for $n > N$,

$$\sum_{i=N+1}^{n} (w_i - W_n' V_n T_i)\, \varepsilon_i = \sum_{i=N+1}^{n} d_i \left\{\varepsilon_i - T_i' V_{i-1}\left(\sum_{j=1}^{i-1} T_j \varepsilon_j\right)\right\}, \tag{2.20}$$

$$s_n = s_N + \sum_{i=N+1}^{n} d_i^2 (1 + T_i' V_{i-1} T_i). \tag{2.21}$$

Since d_n is $\mathscr{F}_{n-1}$-measurable, we obtain by Lemma 2(i) and (2.21) that

$$\sum_{i=1}^{n} d_i \varepsilon_i = O(s_n^{1/2} \max\{1, (\log^+ s_n)^{(1+\delta)/2}\}) \qquad \text{a.s.} \tag{2.22}$$

By the Schwarz inequality, for $n > N$,

$$\begin{aligned} &\left| \sum_{i=N+1}^{n} d_i T_i' V_{i-1} \left(\sum_{j=1}^{i-1} T_j \varepsilon_j \right) \right| \\ &\quad \leqslant \left\{ \sum_{i=N+1}^{n} d_i^2 (1 + T_i' V_{i-1} T_i) \right\}^{1/2} \\ &\quad\quad \times \left\{ \sum_{i=N+1}^{n} \left(T_i' V_{i-1} \sum_{j=1}^{i-1} T_j \varepsilon_j \right)^2 \Big/ (1 + T_i' V_{i-1} T_i) \right\}^{1/2} \\ &\quad \leqslant \left\{ s_n \sum_{i=N+1}^{n} \left[\left(T_i' V_{i-1} \sum_{j=1}^{i-1} T_j \varepsilon_j \right)^2 \Big/ (1 + T_i' V_{i-1} T_i) \right] \right\}^{1/2}, \qquad \text{by (2.21).} \end{aligned} \tag{2.23}$$

By Lemma 1(i, ii), in the event $\{N < \infty\}$,

$$\begin{aligned} &\sum_{i=N+1}^{n} \left(T_i' V_{i-1} \sum_{j=1}^{i-1} T_j \varepsilon_j \right)^2 / (1 + T_i' V_{i-1} T_i) \\ &\qquad = O\left(\max \left\{ 1, \left(\log^+ \lambda_{\max} \left(\sum_{1}^{n} T_i T_i' \right) \right)^{(1+\delta)} \right\} \right) \qquad \text{a.s.} \end{aligned} \tag{2.24}$$

for every $\delta > 0$. From (2.20), (2.22), (2.23) and (2.24), we obtain (2.19). If (1.3) holds, then by Lemma 1(iii) and Lemma 2(ii), (2.22) and (2.24) also hold with $\delta = 0$, and therefore we can set $\delta = 0$ in (2.19). ∎

Proof of Theorem 4. For every nonempty subset J of $\{1,..., p\}$, let $L_{n,J}$ be the linear space spanned by the set of column vectors $\{(z_{1j},..., z_{nj})' : j \in J\}$. Let $T_{n,J}$ denote the column vector whose components are z_{nj}, $j \in J$. (For example, if $J = \{3, 5\}$, then $T_{n,J} = (z_{n3}, z_{n5})'$.) Let $N_J = \inf\{n: \sum_{i=1}^{n} T_{i,J} T_{i,J}'$ is nonsingular$\}$ ($\inf \phi = \infty$). Let $V_{n,J}$ be the Moore–Penrose generalized inverse of $\sum_{i=1}^{n} T_{i,J} T_{i,J}'$, and let $W_{n,J} = \sum_{i=1}^{n} w_i T_{i,J}$. Letting $\hat{\mathbf{w}}_n(J)$ denote the projection of $\mathbf{w}_n$ into $L_{n,J}$, we note that

$$\hat{\mathbf{w}}_n(J) = (T_{1,J},..., T_{n,J})' V_{n,J} W_{n,J}, \tag{2.25}$$

and therefore

$$\langle \mathbf{w}_n(J) - \hat{\mathbf{w}}_n(J), \boldsymbol{\varepsilon}_n \rangle = \sum_{i=1}^{n} \{w_i(J) - W_{n,J}' V_{n,J} T_{i,J}\} \varepsilon_i, \tag{2.26}$$

$$\| \mathbf{w}_n(J) - \hat{\mathbf{w}}_n(J) \|^2 = \sum_{i=1}^{n} \{w_i(J) - W_{n,J}' V_{n,J} T_{i,J}\}^2. \tag{2.27}$$

In view of (2.26) and (2.27), we obtain by Lemma 3 that in the event $\{N_J < \infty\}$,

$$\langle \mathbf{w}_n(J) - \hat{\mathbf{w}}_n(J), \boldsymbol{\varepsilon}_n \rangle = o\left(\| \mathbf{w}_n(J) - \hat{\mathbf{w}}_n(J) \| \times \left[\max \left\{ 1, \log^+ \| \mathbf{w}_n(J) - \hat{\mathbf{w}}_n(J) \|, \log^+ \left(\sum_{j \in J} \sum_{i=1}^{n} z_{ij}^2 \right) \right\} \right]^{(1+\delta)/2} \right) \tag{2.28}$$

a.s. for every $\delta > 0$; moreover, if (1.3) holds, then we can take $\delta = 0$ in (2.28).

Let $\Omega_0 = \{z_{ij} = 0$ for all $i, j\}$. The complementary event is $\Omega_1 = \{\sum_{i=1}^{\infty} z_{ij}^2 > 0$ for some $j\}$. In the event Ω_0, the projection $\hat{\mathbf{w}}_n$ of $\mathbf{w}_n$ into $L(X_n)$ is $\mathbf{0}$ for all n and therefore

$$\begin{aligned}\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle = \langle \mathbf{w}_n, \boldsymbol{\varepsilon}_n \rangle &= \sum_{1}^{n} w_i \varepsilon_i \\ &= o(\| \mathbf{w}_n \| \max\{1, (\log^+ \| \mathbf{w}_n \|)^{(1+\delta)/2}\}) \qquad \text{a.s., by Lemma 2.}\end{aligned}$$

Hence (2.3) holds in the event Ω_0. Now consider the event Ω_1. Obviously for every $\omega \in \Omega_1$, there exists a nonempty subset J (depending on ω) of $\{1, ..., p\}$ such that $N_J < \infty$ and $\hat{\mathbf{w}}_n = \hat{\mathbf{w}}_n(J)$ for all $n \geqslant N_J$ at ω. Since (2.28) holds a.s. for *every* fixed nonempty subset J of $\{1, ..., p\}$ and since there are $2^p - 1$ (and therefore only finitely many) such subsets J, it then follows that $P[(2.28)$ holds for all nonempty subsets J of $\{1, ..., p\}] = 1$, and therefore (2.3) also holds in the event Ω_1. Moreover, if (1.3) holds, then we can take $\delta = 0$ in (2.28) and therefore also in (2.3). ∎

Proof of Theorem 3. Let $c_j(Z_n)$ denote the jth column vector of Z_n and let the $c_j^*(Z_n)$ denote the projection of $c_j(Z_n)$ into the linear space spanned by the vectors $c_1(Z_n), ..., c_{j-1}(Z_n)$, setting $c_1^*(Z_n) = 0$. Since the linear space $L(Z_n)$ is spanned by the orthogonal vectors $c_1(Z_n)$, $c_2(Z_n) - c_2^*(Z_n), ...,$ $c_p(Z_n) - c_p^*(Z_n)$, it then follows that

$$\| \hat{\boldsymbol{\varepsilon}}_n \|^2 = \sum_{j=1}^{p} |\langle c_j(Z_n) - c_j^*(Z_n), \boldsymbol{\varepsilon}_n \rangle|^2 / \| c_j(Z_n) - c_j^*(Z_n) \|^2, \tag{2.29}$$

where the jth summand in (2.29) is taken to be 0 if the denominator $\| c_j(Z_n) - c_j^*(Z_n) \|^2$ is 0. Obviously, $\| c_j(Z_n) - c_j^*(Z_n) \|^2 \leqslant \| c_j(Z_n) \|^2 = \sum_{i=1}^{n} z_{ij}^2$. Therefore by Theorem 4,

$$\begin{aligned}&|\langle c_j(Z_n) - c_j^*(Z_n), \boldsymbol{\varepsilon}_n \rangle|^2 / \| c_j(Z_n) - c_j^*(Z_n) \|^2 \\ &\quad = O\left(\left[\max \left\{ 1, \log^+ \left(\sum_{i=1}^{n} z_{ij}^2 \right), \log^+ \left(\sum_{k=1}^{j-1} \sum_{i=1}^{n} z_{ik}^2 \right) \right\} \right]^{1+\delta} \right) \qquad \text{a.s.}\end{aligned} \tag{2.30}$$

for every $\delta > 0$; moreover, if (1.3) holds, then (2.30) also holds with $\delta = 0$. Applying (2.29) and (2.30), and noting that $\hat{\boldsymbol{\varepsilon}}_n = \mathbf{0}$ if $\sum_{j=1}^{p} \sum_{i=1}^{n} z_{ij}^2 = 0$, we obtain (2.2). ∎

Remark. Under the additional assumption that $P[Z_n' Z_n$ is nonsingular for all large $n] = 1$, Theorem 3 was established by a different method in [10, Lemma 1].

Proof of Theorem 5. Since $\sup_n E(|\varepsilon_n^2 - E(\varepsilon_n^2 | \mathscr{F}_{n-1})|^p | \mathscr{F}_{n-1}) < \infty$ a.s. for some $p > 1$ by (1.3),

$$\sum_1^n \{\varepsilon_i^2 - E(\varepsilon_i^2 | \mathscr{F}_{i-1})\} = o(n) \qquad \text{a.s.} \tag{2.31}$$

(cf. [3]). Since $\sup_i E(\varepsilon_i^2 | \mathscr{F}_{i-1}) < \infty$ a.s. by (1.3), (2.31) implies that $\|\boldsymbol{\varepsilon}_n\|^2 = \sum_1^n \varepsilon_i^2 = O(n)$ a.s. Moreover, since $\liminf_{i \to \infty} E(\varepsilon_i^2 | \mathscr{F}_{i-1}) > 0$ a.s., (2.31) implies that

$$\liminf_{n\to\infty} n^{-1} \|\boldsymbol{\varepsilon}_n\|^2 = \liminf_{n\to\infty} E(\varepsilon_n^2 | \mathscr{F}_{n-1}) > 0 \qquad \text{a.s.} \tag{2.32}$$

By Theorem 3(ii) and (2.4), $\|\hat{\boldsymbol{\varepsilon}}_n\|^2 = o(n)$ a.s. Since $\|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 = \|\boldsymbol{\varepsilon}_n\|^2 - \|\hat{\boldsymbol{\varepsilon}}_n\|^2$, it then follows from (2.32) that $\liminf_{n\to\infty} n^{-1} \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 > 0$ a.s. In view of (2.4), this in turn implies that

$$\max\left\{1, \log^+ \|\mathbf{w}_n - \hat{\mathbf{w}}_n\|, \log^+\left(\sum_{j=1}^{p} \sum_{i=1}^{n} z_{ij}^2\right)\right\} = o(\|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2) \qquad \text{a.s.}$$

Hence by Theorem 4(ii),

$$\langle \mathbf{w}_n - \hat{\mathbf{w}}_n, \boldsymbol{\varepsilon}_n \rangle^2 = o(\|\mathbf{w}_n - \hat{\mathbf{w}}_n\|^2 \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2) \qquad \text{a.s.} \tag{2.33}$$

Moreover, it also follows from Theorem 4(ii) that

$$\begin{aligned} \langle \mathbf{v}_n - \mathbf{v}_n^*, \boldsymbol{\varepsilon}_n \rangle &= o(\|\mathbf{v}_n - \mathbf{v}_n^*\|^2) + o(\|\mathbf{v}_n - \mathbf{v}_n^*\| \, \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|) \\ &= o(\|\mathbf{v}_n - \mathbf{v}_n^*\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2) \qquad \text{a.s.,} \end{aligned} \tag{2.34}$$

$$\langle \mathbf{v}_n - \hat{\mathbf{v}}_n, \boldsymbol{\varepsilon}_n \rangle = o(\|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2) \qquad \text{a.s.} \tag{2.35}$$

Let $\mathbf{r}_n = \mathbf{w}_n - \hat{\mathbf{w}}_n$, and let $\hat{\hat{\mathbf{r}}}_n$ be the projection of $\mathbf{r}_n$ into $L(\mathbf{v}_n + \boldsymbol{\varepsilon}_n)$. Then

$$\mathbf{w}_n - \hat{\mathbf{w}}_n (= \mathbf{r}_n) = \hat{\hat{\mathbf{r}}}_n + (\mathbf{w}_n - \hat{\hat{\mathbf{w}}}_n),$$

and by orthogonality,

$$\|\mathbf{w}_n - \hat{\mathbf{w}}_n\|^2 = \|\hat{\hat{\mathbf{r}}}_n\|^2 + \|\mathbf{w}_n - \hat{\hat{\mathbf{w}}}_n\|^2. \tag{2.36}$$

The projection of $\mathbf{v}_n + \boldsymbol{\varepsilon}_n$ into $L(Z_n)$ is $\hat{\mathbf{v}}_n + \hat{\boldsymbol{\varepsilon}}_n$. Noting that $\hat{\hat{\mathbf{r}}}_n$ is the same as the projection of $\mathbf{r}_n$ (which is orthogonal to $L(Z_n)$) into $L(\mathbf{v}_n + \boldsymbol{\varepsilon}_n - \hat{\mathbf{v}}_n - \hat{\boldsymbol{\varepsilon}}_n)$, we obtain that

$$\|\hat{\hat{\mathbf{r}}}_n\|^2 = |\langle \mathbf{r}_n, \mathbf{v}_n + \boldsymbol{\varepsilon}_n - \hat{\mathbf{v}}_n - \hat{\boldsymbol{\varepsilon}}_n\rangle|^2 / \|\mathbf{v}_n + \boldsymbol{\varepsilon}_n - \hat{\mathbf{v}}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2. \tag{2.37}$$

Likewise, the projection of $\mathbf{v}_n - \hat{\mathbf{v}}_n$ (which is orthogonal to $L(Z_n)$) into $L(Z_n, \mathbf{w}_n)$ is the same as the projection of $\mathbf{v}_n - \hat{\mathbf{v}}_n$ into $L(\mathbf{w}_n - \hat{\mathbf{w}}_n)$ and is therefore of the form $a_n(\mathbf{w}_n - \hat{\mathbf{w}}_n) = a_n \mathbf{r}_n$ for some random variable a_n. Hence

$$\mathbf{v}_n - \hat{\mathbf{v}}_n = a_n \mathbf{r}_n + (\mathbf{v}_n - \mathbf{v}_n^*), \tag{2.38}$$

and therefore by the orthogonality between $\mathbf{r}_n$ and $\mathbf{v}_n - \mathbf{v}_n^*$,

$$\|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^2 = a_n^2 \|\mathbf{r}_n\|^2 + \|\mathbf{v}_n - \mathbf{v}_n^*\|^2, \tag{2.39}$$

$$\langle \mathbf{r}_n, \mathbf{v}_n - \hat{\mathbf{v}}_n\rangle = a_n \|\mathbf{r}_n\|^2. \tag{2.40}$$

Since $\mathbf{r}_n$ is orthogonal to $\hat{\boldsymbol{\varepsilon}}_n$, it follows from (2.40) that

$$\begin{aligned} |\langle \mathbf{r}_n, \mathbf{v}_n - \hat{\mathbf{v}}_n + \boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\rangle|^2 &= (a_n \|\mathbf{r}_n\|^2 + \langle \mathbf{r}_n, \boldsymbol{\varepsilon}_n\rangle)^2 \\ &= a_n^2 \|\mathbf{r}_n\|^4 + 2a_n \langle \mathbf{r}_n, \boldsymbol{\varepsilon}_n\rangle \|\mathbf{r}_n\|^2 + \langle \mathbf{r}_n, \boldsymbol{\varepsilon}_n\rangle^2. \end{aligned} \tag{2.41}$$

By (2.38), (2.39), and the orthogonality between $\mathbf{v}_n - \hat{\mathbf{v}}_n$ and $\hat{\boldsymbol{\varepsilon}}_n$,

$$\begin{aligned} \|(\mathbf{v}_n - \hat{\mathbf{v}}_n) + (\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n)\|^2 &= a_n^2 \|\mathbf{r}_n\|^2 + \|\mathbf{v}_n - \mathbf{v}_n^*\|^2 \\ &\quad + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 + 2a_n \langle \mathbf{r}_n, \boldsymbol{\varepsilon}_n\rangle + 2\langle \mathbf{v}_n - \mathbf{v}_n^*, \boldsymbol{\varepsilon}_n\rangle. \end{aligned} \tag{2.42}$$

From (2.41) and (2.42), it then follows that

$$\begin{aligned} &\|\mathbf{r}_n\|^2 \|\mathbf{v}_n - \hat{\mathbf{v}}_n + \boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 - |\langle \mathbf{r}_n, \mathbf{v}_n + \boldsymbol{\varepsilon}_n - \hat{\mathbf{v}}_n - \hat{\boldsymbol{\varepsilon}}_n\rangle|^2 \\ &\quad = \|\mathbf{r}_n\|^2 \{\|\mathbf{v}_n - \mathbf{v}_n^*\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 + 2\langle \mathbf{v}_n - \mathbf{v}_n^*, \boldsymbol{\varepsilon}_n\rangle\} - \langle \mathbf{r}_n, \boldsymbol{\varepsilon}_n\rangle^2 \\ &\quad = \|\mathbf{r}_n\|^2 (\|\mathbf{v}_n - \mathbf{v}_n^*\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2)(1 + o(1)) \qquad \text{a.s.} \end{aligned} \tag{2.43}$$

The last relation above follows from (2.33) and (2.34). Moreover, by (2.35),

$$\begin{aligned} \|(\mathbf{v}_n - \hat{\mathbf{v}}_n) + (\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n)\|^2 &= \|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 + 2\langle \mathbf{v}_n - \hat{\mathbf{v}}_n, \boldsymbol{\varepsilon}_n\rangle \\ &= (\|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^2 + \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2)(1 + o(1)) \qquad \text{a.s.} \end{aligned} \tag{2.44}$$

From (2.36) and (2.37),

$$\begin{aligned} \|\mathbf{w}_n - \hat{\hat{\mathbf{w}}}_n\|^2 &= \|\mathbf{r}_n\|^2 - \|\hat{\hat{\mathbf{r}}}_n\|^2 \\ &= \frac{\|\mathbf{r}_n\|^2 \|\mathbf{v}_n - \hat{\mathbf{v}}_n + \boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 - |\langle \mathbf{r}_n, \mathbf{v}_n + \boldsymbol{\varepsilon}_n - \hat{\mathbf{v}}_n - \hat{\boldsymbol{\varepsilon}}_n\rangle|^2}{\|\mathbf{v}_n - \hat{\mathbf{v}}_n + \boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2}. \end{aligned} \tag{2.45}$$

Making use of (2.43), (2.44) and (2.45), and noting that $\|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^2 \geqslant \|\mathbf{v}_n - \mathbf{v}_n^*\|^2$, we obtain (2.6). ∎

3. Applications to Autoregressive Processes, Dynamic Systems, and Proof of Theorem 2

We first apply the results of Section 2 to prove Theorem 2 in the following:

Proof of Theorem 2. Without loss of generality we shall only consider the case $j=1$. Relation (1.7) is obvious from the interpretation of $b_{n1}c_1(X_n)+\cdots+b_{np}c_p(X_n)$ as the projection of $\mathbf{y}_n$ into $L(X_n)$. Since $c_1(X_n)-\hat{c}_1(X_n)$ is orthogonal to $c_2(X_n),\dots, c_p(X_n)$ and $\hat{c}_1(X_n)$, it then follows that

$$\begin{aligned}\langle c_1(X_n)-\hat{c}_1(X_n), \mathbf{y}_n\rangle &= \langle c_1(X_n)-\hat{c}_1(X_n), \beta_1 c_1(X_n)+\cdots+\beta_p c_p(X_n)+\boldsymbol{\varepsilon}_n\rangle\\ &=\beta_1\langle c_1(X_n)-\hat{c}_1(X_n), c_1(X_n)\rangle + \langle c_1(X_n)-\hat{c}_1(X_n), \boldsymbol{\varepsilon}_n\rangle\\ &=\beta_1\|c_1(X_n)-\hat{c}_1(X_n)\|^2+\langle c_1(X_n)-\hat{c}_1(X_n), \boldsymbol{\varepsilon}_n\rangle. \end{aligned}\tag{3.1}$$

In view of (1.8) and Theorem 4(ii),

$$\langle c_1(X_n)-\hat{c}_1(X_n), \boldsymbol{\varepsilon}_n\rangle = o(\|c_1(X_n)-\hat{c}_1(X_n)\|^2) \qquad \text{a.s.} \tag{3.2}$$

From (1.7), (3.1) and (3.2), we obtain that $b_{n1}\to\beta_1$ a.s. ∎

Suppose that in the regression model (1.1) the design levels x_{nj} are nonrandom constants and the ε_n are i.i.d. random variables with zero means and unit variances. Then, given $n\geqslant p$, assuming that $X_n'X_n$ is nonsingular and letting $V_n=(v_{ij}^{(n)})_{1\leqslant i,j\leqslant p}=(X_n'X_n)^{-1}$, we have $\operatorname{cov}(b_n)=V_n$, and therefore in particular,

$$v_{jj}^{(n)}=\operatorname{var}(b_{nj})=1/\|c_j(X_n)-\hat{c}_j(X_n)\|^2. \tag{3.3}$$

The last equality above follows from (1.7), noting that X_n is a nonrandom matrix and that $\operatorname{var}(\langle c_j(X_n)-\hat{c}_j(X_n), \mathbf{y}_n\rangle)=E(\langle c_j(X_n)-\hat{c}_j(X_n), \boldsymbol{\varepsilon}_n\rangle)^2$ by (3.1). From (3.3), it then follows that

$$\operatorname{tr}((X_n'X_n)^{-1})=\sum_{j=1}^{p}\|c_j(X_n)-\hat{c}_j(X_n)\|^{-2}. \tag{3.4}$$

The above argument therefore provides a probabilistic proof of the algebraic identity (3.4) for any $n\times p$ matrix X_n of real constants with $n\geqslant p$ and having full rank p. Since $\lambda_{\min}(X_n'X_n)=1/\lambda_{\max}((X_n'X_n)^{-1})$ and since

$\lambda_{\max}((X_n'X_n)^{-1}) \leqslant \operatorname{tr}((X_n'X_n)^{-1}) \leqslant p\lambda_{\max}((X_n'X_n)^{-1})$, we obtain from (3.4) that

$$\begin{aligned} 1/\lambda_{\min}(X_n'X_n) &\leqslant \sum_{j=1}^{p} \|c_j(X_n) - \hat{c}_j(X_n)\|^{-2} \\ &\leqslant p \max_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\|^{-2}, \\ p/\lambda_{\min}(X_n'X_n) &\geqslant \sum_{j=1}^{p} \|c_j(X_n) - \hat{c}_j(X_n)\|^{-2} \\ &\geqslant \max_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\|^{-2}. \end{aligned} \tag{3.5}$$

From (3.5), it then follows that

$$\begin{aligned} &p^{-1} \min_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\|^2 \\ &\qquad \leqslant \lambda_{\min}(X_n'X_n) \leqslant p \min_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\|^2. \end{aligned} \tag{3.6}$$

Since $\lambda_{\min}(X_n'X_n) = 0$ if and only if $\min_{1 \leqslant j \leqslant p} \|c_j(X_n) - \hat{c}_j(X_n)\| = 0$, (3.6) also holds when $X_n'X_n$ is singular. As pointed out in Section 1, the bounds in (3.6) for $\lambda_{\min}(X_n'X_n)$ enable us to analyze the asymptotic properties of $\lambda_{\min}(X_n'X_n)$ by making use of the results in Section 2.

To illustrate these ideas, we first consider the autoregressive AR(p) process

$$y_n = \alpha_1 y_{n-1} + \cdots + \alpha_k y_{n-k} + \varepsilon_n, \qquad n \geqslant 1 \tag{3.7}$$

in the following

COROLLARY 1. *For the autoregressive model* (3.7), *let* $\mathscr{F}_0$ *be the σ-field generated by* $\{y_{1-k}, ..., y_0\}$, *and for* $n \geqslant 1$ *let* $\mathscr{F}_n$ *be the σ-field generated by* $\{y_{1-k}, ..., y_0, \varepsilon_1, ..., \varepsilon_n\}$. *Then* y_n *is* $\mathscr{F}_n$*-measurable. Assume that* $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to* $\{\mathscr{F}_n\}$ *such that* (1.3) *holds and* $\liminf_{n\to\infty} E(\varepsilon_n^2 | \mathscr{F}_{n-1}) > 0$ a.s. *For* $n > k$, *let*

$$X_n = \begin{pmatrix} y_k & \cdots & y_1 \\ y_{k+1} & \cdots & y_2 \\ \vdots & & \\ y_{n-1} & \cdots & y_{n-k} \end{pmatrix} \tag{3.8}$$

be the design matrix for the least-squares estimation of $\alpha_1,\dots,\alpha_k$ *at stage* n. *Suppose that all the roots* z_j *of the characteristic polynomial*

$$\varphi(z) = z^k - \alpha_1 z^{k-1} - \cdots - \alpha_k \tag{3.9}$$

lie on or inside the unit circle (i.e., $|z_j| \leqslant 1$ *for all* $j = 1,\dots,k$*). Then*

$$\liminf_{n\to\infty} n^{-1} \lambda_{\min}(X_n' X_n) > 0 \qquad \text{a.s.} \tag{3.10}$$

and

$$\lambda_{\max}(X_n' X_n) \leqslant \operatorname{tr}(X_n' X_n) = O(n^\gamma) \qquad \text{a.s. } \textit{for some } \gamma \geqslant 1. \tag{3.11}$$

Consequently, the least-squares estimate $(X_n' X_n)^{-1} X_n'(y_{k+1},\dots,y_n)'$ *converges* a.s. *to* $(\alpha_1,\dots,\alpha_k)'$.

We preface the proof of Corollary 1 by the following lemma, which is also used for the proof of Corollaries 2 and 3 below.

LEMMA 4. *Let* $\{\varepsilon_n\}$ *be a martingale difference sequence with respect to an increasing sequence of σ-fields* $\{\mathscr{F}_n\}$ *such that* (2.1) *holds. Suppose that in the dynamic input-output system*

$$y_n = \alpha_1 y_{n-1} + \cdots + \alpha_k y_{n-k} + \boldsymbol{\gamma}_0' \mathbf{u}_n + \cdots + \boldsymbol{\gamma}_h' \mathbf{u}_{n-h} + \varepsilon_n, \tag{3.12}$$

the output y_n *is* $\mathscr{F}_n$*-measurable while the input vector* $\mathbf{u}_n = (u_{n1},\dots,u_{nq})'$ *is* $\mathscr{F}_{n-1}$*-measurable at every stage* n.

(i) *Assume that the roots of the characteristic polynomial* $\varphi(z)$ *as defined in* (3.9) *lie strictly inside the unit circle. Then there exists a positive constant* C *depending only on* $\alpha_1,\dots,\alpha_k,\boldsymbol{\gamma}_0,\dots,\boldsymbol{\gamma}_h$ *such that*

$$\sum_{i=1}^n y_i^2 \leqslant C \left\{ \sum_{i=1}^n \varepsilon_i^2 + \sum_{i=1-h}^n \|\mathbf{u}_i\|^2 + y_0^2 + \cdots + y_{1-k}^2 \right\} \qquad \textit{for all } n \geqslant 1. \tag{3.13}$$

(ii) *Assume that the roots of the characteristic polynomial* $\varphi(z)$ *lie on or inside the unit circle, and that*

$$\|\mathbf{u}_n\|^2 = O(n^\delta) \qquad \text{a.s. } \textit{for some } \delta > 0. \tag{3.14}$$

Then

$$y_n = O(n^a) \qquad \text{a.s. } \textit{for some } a > 0. \tag{3.15}$$

Proof. Letting I_p denote the $p \times p$ identity matrix, define

$$A = \begin{pmatrix} \alpha_1 \cdots \alpha_{k-1} & \alpha_k \\ I_{k-1} & 0 \end{pmatrix}.$$

Viewing A as a linear operator, we define $\|A\| = \sup_{\|x\|=1} \|Ax\|$. To prove (i), since the roots of $\varphi(z)$ lie inside the unit circle, we have

$$\|A^n\| = O(\rho^n) \qquad \text{for some } 0 < \rho < 1, \tag{3.16}$$

as can be easily shown by the Jordan form representation of A (cf. [10]). Define the k-dimensional vectors

$$T_n = (y_n, ..., y_{n-k+1})', \qquad R_n = (\boldsymbol{\gamma}_0' \mathbf{u}_n + \cdots + \boldsymbol{\gamma}_h' \mathbf{u}_{n-h} + \varepsilon_n, 0, ..., 0)'.$$

By (3.12), $T_n = AT_{n-1} + R_n$, and therefore

$$\|T_n\|^2 = \left\| A^n T_0 + \sum_{i=1}^{n} A^{n-i} R_i \right\|^2 \leqslant 2 \|A^n\|^2 \|T_0\|^2 + 2 \left\| \sum_{i=1}^{n} A^{n-i} R_i \right\|^2. \tag{3.17}$$

By the Schwarz inequality,

$$\begin{aligned}
\sum_{n=1}^{N} \left\| \sum_{i=1}^{n} A^{n-i} R_i \right\|^2
&\leqslant \sum_{n=1}^{N} \left(\sum_{i=1}^{n} \|A^{n-i}\| \, \|R_i\| \right)^2 \\
&\leqslant \sum_{n=1}^{N} \left(\sum_{i=1}^{N} \|A^{n-i}\| \right) \left\{ \sum_{i=1}^{n} \|A^{n-i}\| (\boldsymbol{\gamma}_0' \mathbf{u}_i + \cdots \boldsymbol{\gamma}_h' \mathbf{u}_{i-h} + \varepsilon_i)^2 \right\} \\
&\leqslant \left(\sum_{n=0}^{\infty} \|A^n\| \right)^2 \sum_{i=1}^{N} (\boldsymbol{\gamma}_0' \mathbf{u}_i + \cdots + \boldsymbol{\gamma}_h' \mathbf{u}_{i-h} + \varepsilon_i)^2.
\end{aligned} \tag{3.18}$$

From (3.16), (3.17) and (3.18), (3.13) follows. (ii) is established in [10, Theorem 2(i)]. ∎

Proof of Corollary 1. Setting $\mathbf{u}_n = \mathbf{0}$ in (3.12), Eq. (3.12) reduces to (3.7). Since $\operatorname{tr}(X_n' X_n) \leqslant k \sum_{i=1}^{n} y_i^2$, relation (3.11) follows from Lemma 4(ii).

In view of the bounds for $\lambda_{\min}(X_n' X_n)$ in (3.6), it suffices for the proof of (3.10) to show that for $j = 1, ..., p$,

$$\liminf_{n\to\infty} n^{-1} \|c_j(X_n) - \hat{c}_j(X_n)\|^2 > 0 \qquad \text{a.s.} \tag{3.19}$$

Fix j. Let L_j denote the linear space spanned by $\{c_1(X_n), ..., c_{j-1}(X_n)\} \cup \{(y_{k+1-j-r}, ..., y_{n-j-r})' \colon 1 \leqslant r \leqslant k\}$. Let $c_j^*(X_n)$ be the projection of $c_j(X_n)$ into L_j. Since $\hat{c}_j(X_n)$ is the projection of $c_j(X_n)$ into a subspace of L_j,

$\|c_j(X_n)-\hat{c}_j(X_n)\| \geqslant \|c_j(X_n)-c_j^*(X_n)\|$, and therefore to prove (3.19), it suffices to show that

$$\liminf_{n\to\infty} n^{-1}\|c_j(X_n)-c_j^*(X_n)\|^2>0 \quad \text{a.s.} \tag{3.20}$$

We now prove (3.20) by using Theorem 5 and an induction argument. First let Z_n denote the $(n-k)\times k$ matrix whose column vectors are $(y_{k+1-j-r},\ldots, y_{n-j-r})'$, $1\leqslant r\leqslant k$, and let $\pi_{0,n}$ denote the projection of $c_j(X_n)$ into $L(Z_n)$. By (3.7) and (3.8),

$$c_j(X_n)=\sum_{r=1}^{k}\alpha_r(y_{k+1-j-r},\ldots, y_{n-j-r})'+(\varepsilon_{k+1-j},\ldots,\varepsilon_{n-j})'.$$

Therefore, letting $(\tilde{\varepsilon}_{k+1-j},\ldots,\tilde{\varepsilon}_{n-j})'$ denote the projection of $(\varepsilon_{k+1-j},\ldots,\varepsilon_{n-j})'$ into $L(Z_n)$, we have

$$c_j(X_n)-\pi_{0,n}=(\varepsilon_{k+1-j}-\tilde{\varepsilon}_{k+1-j},\ldots,\varepsilon_{n-j}-\tilde{\varepsilon}_{n-j})',$$

and so it follows from Theorem 5 (cf. (2.5)) that

$$\|c_j(X_n)-\pi_{0,n}\|^2=O(n) \quad \text{a.s.}, \qquad \liminf_{n\to\infty} n^{-1}\|c_j(X_n)-\pi_{0,n}\|^2>0 \quad \text{a.s.}, \tag{3.21}$$

noting that $\log \operatorname{tr}(Z_n'Z_n)\leqslant \log(k\sum_1^n y_i^2)=o(n)$ a.s. by Lemma 4(ii).

Since $c_{j-1}(X_n)=\alpha_1 c_j(X_n)+\sum_{r=1}^{k-1}\alpha_{r+1}(y_{k+1-j-r},\ldots, y_{n-j-r})'+\boldsymbol{\varepsilon}_n(1)$, where $\boldsymbol{\varepsilon}_n(1)=(\varepsilon_{k+2-j},\ldots,\varepsilon_{n+1-j})'$, it follows that

$$L(Z_n, c_{j-1}(X_n))=L(Z_n, \alpha_1 c_j(X_n)+\boldsymbol{\varepsilon}_n(1)). \tag{3.22}$$

Now let $\pi_{1,n}$ denote the projection of $c_j(X_n)$ into $L(Z_n, c_{j-1}(X_n))$. In view of (3.22), we can apply Theorem 5 and obtain that

$$\|c_j(X_n)-\pi_{1,n}\|^2=\left\{\frac{\|\boldsymbol{\varepsilon}_n(1)-\hat{\boldsymbol{\varepsilon}}_n(1)\|^2}{\alpha_1^2\|c_j(X_n)-\pi_{0,n}\|^2+\|\boldsymbol{\varepsilon}_n(1)-\hat{\boldsymbol{\varepsilon}}_n(1)\|^2}+o(1)\right\} \times \|c_j(X_n)-\pi_{0,n}\|^2 \quad \text{a.s.}, \tag{3.23}$$

where $\hat{\boldsymbol{\varepsilon}}_n(1)$ is the projection of $\boldsymbol{\varepsilon}_n(1)$ into $L(Z_n)$. Since $\liminf_{n\to\infty} n^{-1}\|\boldsymbol{\varepsilon}_n(1)-\hat{\boldsymbol{\varepsilon}}_n(1)\|^2>0$ a.s. by Theorem 5, it then follows from (3.21) and (3.23) that

$$\|c_j(X_n)-\pi_{1,n}\|^2=O(n) \quad \text{a.s.}, \qquad \liminf_{n\to\infty} n^{-1}\|c_j(X_n)-\pi_{1,n}\|^2>0 \quad \text{a.s.} \tag{3.24}$$

In general, for $2 \leqslant \nu\ (<j)$, we have in analogy with (3.22) that

$$\begin{aligned} &L(Z_n, c_{j-1}(X_n), \ldots, c_{j-\nu}(X_n)) \\ &\quad = L(Z_n, c_{j-1}(X_n), \ldots, c_{j-\nu+1}(X_n), \alpha_\nu c_j(X_n) + \boldsymbol{\varepsilon}_n(\nu)), \end{aligned}$$

where $\boldsymbol{\varepsilon}_n(\nu) = (\varepsilon_{k+\nu+1-j}, \ldots, \varepsilon_{n+\nu-j})'$. Therefore, letting $\pi_{\nu,n}$ denote the projection of $c_j(X_n)$ into $L(Z_n, c_{j-1}(X_n), \ldots, c_{j-\nu}(X_n))$, we can apply Theorem 5 and obtain, in analogy with (3.23), that

$$\begin{aligned} \|c_j(X_n) - \pi_{\nu,n}\|^2 = &\left\{ \frac{\|\boldsymbol{\varepsilon}_n(\nu) - \hat{\boldsymbol{\varepsilon}}_n(\nu)\|^2}{\alpha_\nu^2 \|c_j(X_n) - \pi_{\nu-1,n}\|^2 + \|\boldsymbol{\varepsilon}_n(\nu) - \hat{\boldsymbol{\varepsilon}}(\nu)\|^2} + o(1) \right\} \\ &\times \|c_j(X_n) - \pi_{\nu-1,n}\|^2 \qquad \text{a.s.,} \end{aligned} \tag{3.25}$$

where $\hat{\boldsymbol{\varepsilon}}_n(\nu)$ is the projection of $\boldsymbol{\varepsilon}_n(\nu)$ into $L(Z_n, c_{j-1}(X_n), \ldots, c_{j-\nu+1}(X_n))$. Since $\liminf_{n\to\infty} n^{-1} \|\boldsymbol{\varepsilon}_n(\nu) - \hat{\boldsymbol{\varepsilon}}_n(\nu)\|^2 > 0$ a.s. by Theorem 5, we can apply an induction argument in view of (3.25) to obtain that

$$\|c_j(X_n) - \pi_{\nu,n}\|^2 = O(n) \quad \text{a.s.,} \qquad \liminf_{n\to\infty} n^{-1} \|c_j(X_n) - \pi_{\nu,n}\|^2 > 0 \quad \text{a.s.} \tag{3.26}$$

Since $c_j^*(X_n) = \pi_{j-1,n}$, the desired conclusion (3.20) follows from (3.26) with $\nu = j-1$.

In view of (3.10) and (3.11), we can apply Theorem 1 and obtain the strong consistency of the least-squares estimate in this case. ∎

Corollary 1 generalizes a recent result of Anderson and Taylor [1] on the strong consistency of least-squares estimates for the parameters of a stationary AR(p) process whose characteristic polynomial has all its roots inside the unit circle. In this case,

$$(X_n' X_n)/n \to \Gamma \qquad \text{a.s. for some nonrandom positive definite symmetric matrix } \Gamma, \tag{3.27}$$

and therefore the eigenvalues of $X_n' X_n$ are all of linear order n. By allowing the roots to lie on the unit circle as well, we can formulate nonstationary models related to the ARIMA models of Box and Jenkins [2]. In this nonstationary case, the asymptotic behavior of $X_n' X_n$ is much more complex than (3.27). Recently, Moore [13, p. 506] asserted that the eigenvalues of $X_n' X_n$ are of quadratic order n^2 when the roots of $\varphi(z)$ are all on the unit circle. This turns out to be false as shown by the following

EXAMPLE 3. Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with mean 0 and variance $\sigma^2 > 0$. Consider the AR(1) process

$$y_n = y_{n-1} + \varepsilon_n, \qquad n \geqslant 1.$$

Thus, $y_n = y_0 + S_n$, where $S_n = \sum_1^n \varepsilon_i$. By Strassen's law of the iterated logarithm [14],

$$\limsup_{n\to\infty} \left(\sum_1^n S_i^2\right)/(n^2 \log\log n) = 8\sigma^2/\pi^2 \qquad \text{a.s.} \tag{3.28}$$

On the other hand, by a result of Donsker and Varadhan [5, p. 751],

$$\liminf_{n\to\infty} \left(\sum_1^n S_i^2\right)/\{n^2/\log\log n\} = \sigma^2/4 \qquad \text{a.s.} \tag{3.29}$$

Hence, with probability 1, $X_n' X_n = \sum_1^{n-1} y_i^2 \sim \sum_1^{n-1} S_i^2$ fluctuates between $(\frac{1}{4} + o(1))\,\sigma^2 n^2/\log\log n$ and $(8\pi^{-2} + o(1))\,\sigma^2 n^2 \log\log n$ in this case.

The following example shows that in Corollary 1 it is possible for $\lambda_{\min}(X_n' X_n)$ to have linear order n and $\lambda_{\max}(X_n' X_n)$ to attain a higher algebraic order.

EXAMPLE 4. Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with mean 0 and variance $\sigma^2 > 0$. Consider the AR(2) process

$$y_n = \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \varepsilon_n, \qquad n \geqslant 1,\ y_0 = y_{-1} = 0.$$

Then

$$X_n' X_n = \begin{pmatrix} \sum_2^{n-1} y_i^2 & \sum_2^{n-1} y_i y_{i-1} \\ \sum_2^{n-1} y_i y_{i-1} & \sum_1^{n-2} y_i^2. \end{pmatrix}. \tag{3.30}$$

Assume that $\alpha_1 = 1$ and $\alpha_2 = 0$. Then $y_n = S_n (= \sum_1^n \varepsilon_i)$, and the characteristic polynomial $\varphi(z) = z^2 - z$ has 1 and 0 as its roots. Since $y_i = S_i$, it follows from (3.30) that

$$\lambda_{\max}(X_n' X_n) + \lambda_{\min}(X_n' X_n) = \sum_2^{n-1} S_i^2 + \sum_1^{n-2} S_i^2, \tag{3.31}$$

$$\begin{aligned} \lambda_{\max}(X_n' X_n)\,\lambda_{\min}(X_n' X_n) &= \left(\sum_2^{n-1} S_i^2\right)\left(\sum_2^{n-1} S_{i-1}^2\right) - \left(\sum_2^{n-1} S_i S_{i-1}\right)^2 \\ &= \left(\sum_2^{n-1} \varepsilon_i^2\right)\left(\sum_2^{n-1} S_{i-1}^2\right) - \left(\sum_2^{n-1} \varepsilon_i S_{i-1}\right)^2. \end{aligned} \tag{3.32}$$

Since $(\sum_2^{n-1} \varepsilon_i S_{i-1})^2 = o((\sum_2^{n-1} S_{i-1}^2)(\log n)^{1+\delta})$ a.s. for all $\delta > 0$ by Lemma 2 and (3.28) and since $\sum_2^{n-1} \varepsilon_i^2 \sim n\sigma^2$ a.s., it then follows from (3.31) and (3.32) that

$$\lambda_{\max}(X_n' X_n) \sim 2 \sum_1^n S_i^2 \qquad \text{a.s.}, \qquad \lambda_{\min}(X_n' X_n) \sim \tfrac{1}{2} n\sigma^2 \qquad \text{a.s.}, \tag{3.33}$$

and therefore by (3.28) and (3.29), $\lambda_{\max}(X_n' X_n)$ oscillates a.s. between $(\frac{1}{2} + o(1))\,\sigma^2 n^2/\log\log n$ and $(16\pi^{-2} + o(1))\,\sigma^2 n^2 \log\log n$.

We now apply Theorems 1 and 5 to extend a recent theorem of Christopeit and Helmes [4] on the strong consistency of least-squares estimates for the parameters of the dynamic system (3.12). Assuming the roots of the characteristic polynomial $\varphi(z)$ as defined in (3.9) to lie strictly inside the unit circle, Christopeit and Helmes [4] considered the case of nonrandom inputs $\mathbf{u}_n$ (exogenous variables) with $h = 0$ in (3.12) under the assumptions

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_{i+\nu}' \qquad \text{exists for every } \nu = 0, 1, 2, \ldots, \tag{3.34a}$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i' \qquad \text{is positive definite.} \tag{3.34b}$$

By making use of Theorem 5, we can remove assumption (3.34a) together with the requirement that $\mathbf{u}_n$ be nonstochastic and replace assumption (3.34b) by the weaker assumption (3.35) below. This is the content of

COROLLARY 2. *For the dynamic input–output system* (3.12), *let* $\mathscr{F}_n$ *be the* σ*-field generated by* $\{y_{1-k}, \ldots, y_0, \varepsilon_1, \ldots, \varepsilon_n\} \cup \{\mathbf{u}_{1-h}, \ldots, \mathbf{u}_{n+k+1}\}$ *for* $n \geq 1$. *Then at every stage n, the output* y_n *is* $\mathscr{F}_n$*-measurable while the input vector* $\mathbf{u}_n = (u_{n1}, \ldots, u_{nq})'$ *is* $\mathscr{F}_{n-k-1}$*-measurable. Assume that* $\{\varepsilon_n\}$ *is a martingale difference sequence with respect to* $\{\mathscr{F}_n\}$ *such that* (1.3) *holds and* $\liminf_{n\to\infty} E(\varepsilon_n^2 | \mathscr{F}_{n-1}) > 0$ a.s. *Suppose that the roots* z_j *of the characteristic polynomial* $\varphi(z)$ *defined in* (3.9) *lie inside the unit circle* (*i.e.*, $|z_j| < 1$ *for all* $j = 1, \ldots, k$). *Letting* $U_n = (\mathbf{u}_n', \mathbf{u}_{n-1}', \ldots, \mathbf{u}_{n-h}')'$, *assume that*

$$\operatorname{tr}\left(\sum_{i=h+1}^n U_i U_i'\right) = O(n) \quad \text{a.s.}, \quad \liminf_{n\to\infty} n^{-1} \lambda_{\min}\left(\sum_{i=h+1}^n U_i U_i'\right) > 0 \quad \text{a.s.} \tag{3.35}$$

For $n \geq I > \max(h, k)$, *let*

$$X_n = \begin{pmatrix} y_{I-1} & \cdots & y_{I-k} & \mathbf{u}_I' & \cdots & \mathbf{u}_{I-h}' \\ \vdots & & & & & \\ y_{n-1} & \cdots & y_{n-k} & \mathbf{u}_n' & \cdots & \mathbf{u}_{n-h}' \end{pmatrix} \tag{3.36}$$

be the design matrix for the least-squares estimation of $\alpha_1,\ldots,\alpha_k,\gamma_0,\ldots,\gamma_h$ *at stage* n. *Then*

$$\operatorname{tr}(X_n'X_n)=O(n)\quad\text{a.s.},\quad \liminf_{n\to\infty} n^{-1}\lambda_{\min}(X_n'X_n)>0\quad\text{a.s.}\tag{3.37}$$

Consequently, the least-squares estimate $(X_n'X_n)^{-1}X_n'(y_I,\ldots,y_n)'$ *converges a.s. to* $(\alpha_1,\ldots,\alpha_k,\gamma_0',\ldots,\gamma_h')'$.

Proof. Since $\sum_1^n \varepsilon_i^2=O(n)$ a.s. by (2.5) and $\sum_1^n \|\mathbf{u}_i\|^2=O(n)$ a.s. by (3.35), it follows from Lemma 4(i) and (3.35) that

$$\operatorname{tr}(X_n'X_n)\leqslant k\sum_1^n y_i^2+\operatorname{tr}\left(\sum_{i=h+1}^n U_iU_i'\right)=O(n)\qquad\text{a.s.}\tag{3.38}$$

We now apply (3.6) and Theorem 5 to prove that $\liminf_{n\to\infty} n^{-1}\lambda_{\min}(X_n'X_n)>0$ a.s. First note that $X_n=(\mathbf{y}_n(1),\ldots,\mathbf{y}_n(k),W_n)$, where

$$\mathbf{y}_n(v)=(y_{I-v},\ldots,y_{n-v})',\qquad v=0,1,2,\ldots;\tag{3.39}$$

$$W_n=\begin{pmatrix} u_{I1} & \cdots & u_{Iq} & \cdots & u_{I-h,1} & \cdots & u_{I-h,q}\\ \vdots & & & & & & \\ u_{n1} & \cdots & u_{nq} & \cdots & u_{n-h,1} & \cdots & u_{n-h,q}\end{pmatrix}=\sum_{i=I}^n U_iU_i'.\tag{3.40}$$

Take a column vector $\mathbf{w}_n=(w_I,\ldots,w_n)'$ of W_n and let W_n^* denote the submatrix consisting of all the other column vectors of W_n. Letting $\hat{\mathbf{w}}_n$ be the projection of $\mathbf{w}_n$ into $L(W_n^*)$, we obtain by (3.6) and (3.35) that

$$\liminf_{n\to\infty} n^{-1}\|\mathbf{w}_n-\hat{\mathbf{w}}_n\|^2>0\qquad\text{a.s.}\tag{3.41}$$

Now let $\hat{\mathbf{w}}_{n,k}$ be the projection of $\mathbf{w}_n$ into $L(W_n^*,\mathbf{y}_n(k))$. By (3.12) and the fact that $\mathbf{u}_n$ is $\mathscr{F}_{n-k-1}$-measurable,

$$\mathbf{y}_n(k)=(y_{I-k},\ldots,y_{n-k})'=(v_I,\ldots,v_n)'+(\varepsilon_{I-k},\ldots,\varepsilon_{n-k})',\tag{3.42}$$

where $v_i=\alpha_1 y_{i-k-1}+\cdots+\alpha_k y_{i-2k}+\gamma_0'\mathbf{u}_{i-k}+\cdots+\gamma_h'\mathbf{u}_{i-k-h}$ is $\mathscr{F}_{i-k-1}$-measurable; moreover, $\mathbf{u}_i$ and therefore w_i also are $\mathscr{F}_{i-k-1}$-measurable. Hence, by Theorem 5,

$$\|\mathbf{w}_n-\hat{\mathbf{w}}_{n,k}\|^2\geqslant\left\{\frac{\|\boldsymbol{\varepsilon}_n-\hat{\boldsymbol{\varepsilon}}_n\|^2}{\|\mathbf{y}_n(k)-\hat{\mathbf{y}}_n(k)\|^2+\|\boldsymbol{\varepsilon}_n-\hat{\boldsymbol{\varepsilon}}_n\|^2}+o(1)\right\}\|\mathbf{w}_n-\hat{\mathbf{w}}_n\|^2\qquad\text{a.s.,}\tag{3.43}$$

where $\boldsymbol{\varepsilon}_n = (\varepsilon_{I-k},..., \varepsilon_{n-k})'$, and $\hat{\mathbf{y}}_n(k)$, $\hat{\boldsymbol{\varepsilon}}_n$ are respectively the projections of $\mathbf{y}_n(k)$, $\boldsymbol{\varepsilon}_n$ into $L(W_n^*)$. Since $\|\mathbf{y}_n(k)\|^2 = O(n)$ a.s. by (3.38) and $\liminf_{n\to\infty} n^{-1} \|\boldsymbol{\varepsilon}_n - \hat{\boldsymbol{\varepsilon}}_n\|^2 > 0$ a.s. by Theorem 5, it follows from (3.41) and (3.42) that

$$\liminf_{n\to\infty} n^{-1} \|\mathbf{w}_n - \hat{\mathbf{w}}_{n,k}\|^2 > 0 \qquad \text{a.s.} \tag{3.44}$$

Now let $\hat{\mathbf{w}}_{n,k-1}$ be the projection of $\mathbf{w}_n$ into $L(W_n^*, \mathbf{y}_n(k), \mathbf{y}_n(k-1))$. By (3.44) and a similar argument as in (3.43), we then obtain that $\liminf_{n\to\infty} n^{-1} \|\mathbf{w}_n - \hat{\mathbf{w}}_{n,k-1}\|^2 > 0$ a.s. Proceeding inductively in this way, we finally obtain that

$$\liminf_{n\to\infty} n^{-1} \|\mathbf{w}_n - \hat{\mathbf{w}}_{n,1}\|^2 > 0 \qquad \text{a.s.,} \tag{3.45}$$

where $\hat{\mathbf{w}}_{n,1}$ is the projection of $\mathbf{w}_n$ into $L(W_n^*, \mathbf{y}_n(k),..., \mathbf{y}_n(1))$.

Now take the column vector $\mathbf{y}_n(v)$ $(v = 1,..., k)$ of X_n and let $\mathbf{y}_n^*(v)$ be its projection into the linear space generated by the other column vectors of X_n. By Corollary 3(i) below, $\liminf_{n\to\infty} n^{-1} \|\mathbf{y}_n(v) - \mathbf{y}_n^*(v)\|^2 > 0$ a.s. This and (3.45) in turn imply that $\liminf_{n\to\infty} n^{-1} \lambda_{\min}(X_n' X_n) > 0$ a.s. by (3.6). ∎

The following corollary considers dynamic systems (3.12) whose characteristic polynomials may have roots lying on the unit circle. While relation (3.38), which is crucial in the preceding argument, may no longer hold in this more general case, we can modify the preceding proof and combine it with the ideas in Corollary 1 to obtain a partial extension of Corollary 2 in the following:

COROLLARY 3. *Suppose that for the dynamic input–output system* (3.12), *the roots z_j of the characteristic polynomial $\varphi(z)$ as defined in* (3.9) *lie on or inside the unit circle (i.e., $|z_j| \leqslant 1$ for $j = 1,..., k$). Assume that the input vectors $\mathbf{u}_n = (u_{n1},..., u_{nq})'$ satisfy* (3.14). *Let $\mathscr{F}_n$ be the σ-field generated by $\{y_{1-k},..., y_0, \varepsilon_1,..., \varepsilon_n\} \cup \{\mathbf{u}_{1-h},..., \mathbf{u}_{n+k+1}\}$, and assume that $\{\varepsilon_n\}$ is a martingale difference sequence with respect to $\{\mathscr{F}_n\}$ such that* (1.3) *holds and* $\liminf_{n\to\infty} E(\varepsilon_n^2 | \mathscr{F}_{n-1}) > 0$ a.s. *For $n \geqslant I > \max(h, k)$, define the design matrix X_n as in* (3.36) *and also define the column vectors $\mathbf{y}_n(v)$ as in* (3.39).

(i) *For fixed $v = 1,..., k$, let X_n^* be the submatrix of X_n with the column vector $\mathbf{y}_n(v)$ removed, and let $\mathbf{y}_n^*(v)$ denote the projection of $\mathbf{y}_n(v)$ into $L(X_n^*)$. Then*

$$\liminf_{n\to\infty} n^{-1} \|\mathbf{y}_n(v) - \mathbf{y}_n^*(v)\|^2 > 0 \qquad \text{a.s.} \tag{3.46}$$

Consequently the least-squares estimate

$$\hat{\alpha}_n(v) = \langle \mathbf{y}_n(v) - \mathbf{y}_n^*(v), \mathbf{y}_n(0) \rangle / \| \mathbf{y}_n(v) - \mathbf{y}_n^*(v) \|^2 \tag{3.47}$$

converges a.s. *to* α_v.

(ii) *For fixed* $v = 0,..., h$ *and* $j = 1,..., q$, *the least-squares estimate* $\hat{\gamma}_n(v,j)$ *at stage n of the component* γ_{vj} *in the vector* $\boldsymbol{\gamma}_v = (\gamma_{v1},..., \gamma_{vq})'$ *is given by*

$$\hat{\gamma}_n(v,j) = \langle \mathbf{w}_n(v,j) - \hat{\mathbf{w}}_n(v,j), \mathbf{y}_n(0) \rangle / \| \mathbf{w}_n(v,j) - \hat{\mathbf{w}}_n(v,j) \|^2, \tag{3.48}$$

where $\mathbf{w}_n(v,j) = (u_{l-v,j},..., u_{n-v,j})'$ *and* $\hat{\mathbf{w}}_n(v,j)$ *is the projection of* $\mathbf{w}_n(v,j)$ *into* $L(\hat{X}_n)$, $\hat{X}_n$ *being the submatrix of* X_n *with the column vector* $\mathbf{w}_n(v,j)$ *removed. Fix* $j\ (= 1,..., q)$. *Assume furthermore that the inputs* u_{nj}, $n \geqslant 1$, *are independent random variables satisfying*

$$Eu_{nj} = 0 \quad \text{for all } n \geqslant 1, \quad \liminf_{n \to \infty} Eu_{nj}^2 > 0, \quad \text{and} \quad \sup_{n \geqslant 1} E|u_{nj}|^\alpha < \infty \tag{3.49}$$

for some $\alpha > 2$, *and are independent of the set of random variables* $\{y_{1-k},..., y_0;\ \mathbf{u}_{1-h},..., \mathbf{u}_0;\ \varepsilon_1, \varepsilon_2,...\} \cup \{u_{ni} : i \neq j, n \geqslant 1\}$. *Then for every* $v = 0,..., h$,

$$\liminf_{n \to \infty} n^{-1} \| \mathbf{w}_n(v,j) - \hat{\mathbf{w}}_n(v,j) \|^2 > 0 \qquad \text{a.s.}, \tag{3.50}$$

and consequently $\hat{\gamma}_n(v,j) \to \gamma_{vj}$ a.s.

Remark. Let $\mathscr{G}_n$ be the σ-field generated by $\{y_{1-k},..., y_0, \mathbf{u}_{1-h},..., \mathbf{u}_0;\ \varepsilon_1,..., \varepsilon_n\}$ and assume that $\{\varepsilon_n\}$ is a martingale difference sequence with respect to $\{\mathscr{G}_n\}$ such that $\sup_n E(|\varepsilon_n|^\alpha \mid \mathscr{G}_{n-1}) < \infty$ a.s. for some $\alpha > 2$ and $\liminf_{n\to\infty} E(\varepsilon_n^2 \mid \mathscr{G}_{n-1}) > 0$. Consider the dynamic system (3.12) with errors ε_n and (white noise) inputs u_{nj} $(n \geqslant 1,\ j = 1,..., q)$ which are independent random variables satisfying (3.49) for every j and are independent of the initial values $y_{1-k},..., y_0, \mathbf{u}_{1-h},..., \mathbf{u}_0$ and the noise sequence $\{\varepsilon_n\}$. Then defining $\mathscr{F}_n$ as in Corollary 3, $\{\varepsilon_n\}$ is also a martingale difference sequence with respect to $\{\mathscr{F}_n\}$. Furthermore, (3.49) implies that (3.14) is satisfied. Hence, by Corollary 3 and (3.6), $\liminf_{n\to\infty} n^{-1}\lambda_{\min}(X_n' X_n) > 0$ a.s.; moreover, the least-squares estimates of $\alpha_1,..., \alpha_k, \boldsymbol{\gamma}_0,..., \boldsymbol{\gamma}_h$ are strongly consistent.

Proof of Corollary 3. To prove (i), let

$$X_n(v) = \left(\mathbf{y}_n(v+1),..., \mathbf{y}_n(v+k) \ \vdots \ \begin{matrix} \mathbf{u}_l' & \cdots & \mathbf{u}_{l-v-h}' \\ \vdots & & \\ \mathbf{u}_n' & \cdots & \mathbf{u}_{n-v-h} \end{matrix} \right)$$

and let L_ν be the linear space spanned by $\mathbf{y}_n(1),..., \mathbf{y}_n(\nu-1)$ and the column vectors of $X_n(\nu)$. Let $\hat{\mathbf{y}}_n(\nu)$ be the projection of $\mathbf{y}_n(\nu)$ into L_ν. Since $\mathbf{y}_n^*(\nu)$ is the projection of $\mathbf{y}_n(\nu)$ into a subspace of L_ν, $\|y_n(\nu)-\mathbf{y}_n^*(\nu)\| \geqslant \|\mathbf{y}_n(\nu)-\hat{\mathbf{y}}_n(\nu)\|$, and therefore to prove (3.46), it suffices to show that

$$\liminf_{n\to\infty} n^{-1} \|\mathbf{y}_n(\nu)-\hat{\mathbf{y}}_n(\nu)\|^2 > 0 \qquad \text{a.s.} \tag{3.51}$$

In view of assumption (3.14), we can apply Lemma 4(ii) to conclude that (3.15) holds and therefore $\operatorname{tr}\{X_n'(\nu)\, X_n(\nu)\} = O(n^\rho)$ for some $\rho > 1$. By (3.12), $\mathbf{y}_n(\nu) = (y_{I-\nu},..., y_{n-\nu})'$ is a linear combination of the column vectors of $X_n(\nu)$ plus the vector $\boldsymbol{\varepsilon}_n = (\varepsilon_{I-\nu},..., \varepsilon_{n-\nu})'$. Therefore, letting $\pi_{0,n}$ be the projection of $\mathbf{y}_n(\nu)$ into $L(X_n(\nu))$ and noting that $\mathbf{y}_n(\nu+1)$ and $\mathbf{u}_n'$ are $\mathscr{F}_{n-\nu-1}$-measurable, we obtain from Theorem 5 by a similar argument as in Corollary 1 (cf. (3.21)) that for $i=0$,

$$\|\mathbf{y}_n(\nu)-\pi_{i,n}\|^2 = O(n) \quad \text{a.s.,} \quad \liminf_{n\to\infty} n^{-1} \|\mathbf{y}_n(\nu)-\pi_{i,n}\|^2 > 0 \quad \text{a.s.} \tag{3.52}$$

Now let $\pi_{1,n}$ be the projection of $\mathbf{y}_n(\nu)$ into $L(X_n(\nu), \mathbf{y}_n(\nu-1))$ and note that as in (3.22),

$$L(X_n(\nu), \mathbf{y}_n(\nu-1)) = L(X_n(\nu), \alpha_1 \mathbf{y}_n(\nu) + \boldsymbol{\varepsilon}_n(1)),$$

where $\boldsymbol{\varepsilon}_n(1) = (\varepsilon_{I-\nu+1},..., \varepsilon_{n-\nu+1})'$. Hence by a similar argument as in Corollary 1, it then follows that (3.50) also holds with $i=1$. In general, letting $\pi_{i,n}$ be the projection of $\mathbf{y}_n(\nu)$ into $L(X_n(\nu), \mathbf{y}_n(\nu-1),..., \mathbf{y}_n(\nu-i))$, an induction argument like that in Corollary 1 then shows that (3.52) holds for $i=1,..., \nu-1$. Since $\hat{\mathbf{y}}_n(\nu) = \pi_{\nu-1,n}$, we then obtain the desired conclusion (3.51), and therefore (3.46) holds. The representation (3.47) for $\hat{\alpha}_n(\nu)$ is a special case of (1.7), and the strong consistency of $\hat{\alpha}_n(\nu)$ follows from Theorem 2.

The proof of (ii) is completely analogous. Fix $\nu\,(=0,..., h)$ and j $(=1,..., q)$, and let $W_n(\nu,j)$ be the matrix whose column vectors are

$$\begin{aligned} &\mathbf{y}_n(\nu+1),..., \mathbf{y}_n(\nu+k); \qquad && \mathbf{w}_n(\nu+1,j),..., \mathbf{w}_n(\nu+h,j); \\ &\mathbf{w}_n(0,i),..., \mathbf{w}_n(\nu+h,i), \qquad && i\neq j \text{ and } 1\leqslant i\leqslant q. \end{aligned} \tag{3.53}$$

Note that $\mathbf{w}_n(\nu,j) = (u_{I-\nu,j},..., u_{n-\nu,j})'$ is independent of $\{\varepsilon_n : n\geqslant 1\} \cup \{\mathbf{w}_n(m,i) : 0\leqslant m\leqslant \nu+h, i\neq j\}$. We now outline the stepwise argument for the proof of (3.50). First, project $\mathbf{w}_n(\nu,j)$ into $L(W_n(\nu,j))$ and apply Theorem 5 as before. Then project $\mathbf{w}_n(\nu,j)$ into

$$L(W_n(\nu,j), \mathbf{y}_n(\nu)) = L(W_n(\nu,j), \gamma_{0j}\mathbf{w}_n(\nu,j) + \boldsymbol{\varepsilon}_n),$$

where $\boldsymbol{\varepsilon}_n = (\varepsilon_{l-\nu}, \ldots, \varepsilon_{n-\nu})'$, and then project it into $L(W_n(\nu, j)$, $\mathbf{y}_n(\nu)$, $\mathbf{w}_n(\nu - 1, j))$. Proceeding inductively in this way and applying Theorem 5 at each step, we finally obtain as before that $\liminf_{n\to\infty} n^{-1} \|\mathbf{w}_n(\nu,j) - \hat{\hat{\mathbf{w}}}_n(\nu,j)\|^2 > 0$ a.s., where $\hat{\hat{\mathbf{w}}}_n(\nu,j)$ is the projection of $\mathbf{w}_n(\nu,j)$ into $L(W_n(\nu,j);$ $\mathbf{y}_n(\nu)$, $\mathbf{w}_n(\nu-1,j); \ldots; \mathbf{y}_n(1), \mathbf{w}_n(0,j))$, which contains $L(\hat{X}_n)$ as a subspace. Since $\|\mathbf{w}_n(\nu,j) - \hat{\mathbf{w}}_n(\nu,j)\| \geqslant \|\mathbf{w}_n(\nu,j) - \hat{\hat{\mathbf{w}}}_n(\nu,j)\|$, the desired conclusion (3.50) follows. ∎

References

[1] Anderson, T. W., and Taylor, J. B. (1979). Strong consistency of least squares estimators in dynamic models. *Ann. Statist.* **7** 484–489.

[2] Box, G. E. P., and Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control.* Holden Day, San Francisco.

[3] Chow, Y. S. (1965). Local convergence of martingales and the law of large numbers. *Ann. Math. Statist.* **36** 552–558.

[4] Christopeit, N., and Helmes, K. (1980). Strong consistency of least squares estimators in linear regression models. *Ann. Statist.* **8** 778–788.

[5] Donsker, M. D., and Varadhan, S. R. S. (1977). On laws of the iterated logarithm for local times. *Comm. Pure Appl. Math.* **30** 707–753.

[6] Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrsch. Verw. Gebiete* **34** 119–127.

[7] Lai, T. L., and Robbins, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

[8] Lai, T. L., and Robbins, H. (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Z. Wahrsch. Verw. Gebiete* **56** 329–360.

[9] Lai, T. L., Robbins, H., and Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9** 343–361.

[10] Lai, T. L., and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166.

[11] Ljung, L. (1976). Consistency of the least squares identification method. *IEEE Trans. Automat. Control* **AC-21** 779–781.

[12] Ljung, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Automat. Control* **AC-21** 551–575.

[13] Moore, J. B. (1978). On strong consistency of least squares identification algorithms. *Automatica* **14** 505–509.

[14] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Gebiete* **3** 211–226.